Pattern Recognition Options to Combine Process Monitoring and Material Accounting Data in Nuclear Safeguards

[1]T. Burr, [2]M.S. Hamada, [3]M. Skurikhin, [4]B. Weaver

Statistical Sciences, Los Alamos National Laboratory, United States

[1]tburr@lanl.gov; [2]hamada@lanl.gov; [3]misha@lanl.gov; [4]theguz@lanl.gov

Abstract

Nuclear material accounting (NMA) is a component of nuclear safeguards, which are designed to deter and detect illicit diversion of special nuclear material (SNM) from the peaceful fuel cycle to a weapons program. NMA consists of periodically, but at relatively low frequency, comparing measured SNM inputs to measured SNM outputs, and adjusting for measured changes in inventory. Process monitoring (PM) is a relatively recent component of safeguards that consists of data more frequently collected than NMA data. PM data are often only an indirect measurement of the SNM and is typically used as a qualitative measure to supplement NMA, or to support indirect estimation of difficult-to-measure inventory for NMA. This paper introduces quantitative diversion detection options for NMA and PM data, which can be regarded as time series of residuals. Unique statistical challenges in combining NMA and PM residual time series include: PM and NMA data are collected at different frequencies; PM residuals often have a probability distribution that cannot be adequately modeled by a Gaussian distribution, not all PM and NMA data streams are independent, and the monitoring scheme must have reasonably high detection probability for both abrupt and protracted diversion.

Keywords

Data-driven Hypothesis Testing; Mixture Distribution; Pattern Recognition; Sequential Testing

Introduction

In traditional nuclear safeguards, periodic nuclear materials accounting (NMA) measurements aim to confirm the presence of special nuclear material (SNM) in accountability vessels to within relatively small tolerances arising from measurement errors. Traditional NMA at large throughput facilities closes the material balance (MB) approximately every 10 to 30 days around an entire material balance area, which typically consists of multiple process stages. The example facility used in this paper is an aqueous reprocessing facility [1], which often has large throughput and many tens of tanks plus various types of processing equipment.

The MB is defined as MB = $T_{in} + I_{begin} - T_{out} - I_{end}$, where T_{in} is transfers in, T_{out} is transfers out, I_{begin} is beginning inventory, and I_{end} is ending inventory. The measurement error standard deviation of the MB is denoted σ_{MB}. For large throughput nuclear facilities, such as commercial reprocessing plants, it is difficult to satisfy NMA goals for detecting diversion. Therefore, additional measures are taken to supplement NMA. One additional measure is process monitoring (PM) [2-5], which has recognized but currently unquantified benefits. Despite occasional attempts to quantify the diversion detection capability of PM, quantitative claims regarding safeguards effectiveness involve NMA, with PM regarded as a qualitative added measure or used in a support-to-NMA role. A common support-to-NMA role is for PM to help provide estimates of difficult-to-measure in-process inventory.

There are many roles for PM [4], and PM data come in a variety of forms [4, 5]. PM often involves more frequent but lower quality measurements than NMA [4]. While NMA estimates SNM mass balances and uncertainties, PM sometimes tracks SNM attributes qualitatively, or in the case of solution monitoring, might track bulk mass and volume. PM data can also include very frequent high-dimensional spectral data from gamma detectors [6], or low-dimensional flow and/or in-tank volume data from flow meters or in-tank dip tubes. In some cases, PM

data can be relatively high quality, such as in-line mass flow measurements, and some current research is aimed at high-quality in-line SNM accountability measurements [7].

PM includes analyzing process control measurements to detect abnormal plant operation. Process control measurements are those used by the operator to control the chemical and/or physical processes. Example process control measurements in an aqueous reprocessing plant include (1) temperature, mass, and density measurements in tanks, (2) inline flow meters, (3) concentration measurement of nonnuclear materials. Figure 1 is a diagram of a generic aqueous facility as modeled in [8].

This paper focuses on PM in which in-tank measurements of bulk mass and/or Pu mass are available. However, if PM is viewed as a type of modern near-real-time accounting (NRTA), [9] showed that protracted diversion detection is still very difficult. Therefore, we introduce a new concept of a model-based prediction for each SNM flow stream so that time series of residuals can monitor for diversion from any given stream.

The work described here illustrates options to quantify the benefit of using both PM and NMA data on the same footing, defining the system alarm probability P(alarm | diversion scenario), as the conditional probability of an alarm given the true model parameters (such as the true SNM loss in each tank over a specified time). The estimated model parameters lead to p residuals r_1, r_2, …., r_p, which include recent MBs from NMA, plus residuals generated from PM data such as, for example, solution levels in tanks. The probability P(alarm | diversion scenario) is a function of the true states of nature which depend on whether SNM has been misdirected, the measurement system, and the alarm rule(s).

The following sections include related work, a description of NMA and of PM, event marking, data fusion, pattern recognition, model-based prediction, examples using a 2-tank and a 7-tank material balance area (MBA), discussion of available simulated and real data, extensions to include additional PM data, and summary. Appendix 1 provides a flow chart of the 5 main analysis steps used in the two main examples in Section IX.

Related Work

This section reviews related work in the nuclear safeguards and statistics communities.

A. *Related Work in Nuclear Safeguards for NMA and PM*

The use of PM data for safeguards dates back to at least the 1980s when the Barnwell reprocessing plant included unit process accounting areas such as individual tanks, and NMA was performed daily or on a data-driven basis, as in "near-real-time" accounting (NRTA) [10]. More recently, solution monitoring (SM) as an example of PM is being used to complement NMA [4, 5, 8, 11].

The only other attempt the authors are aware of to quantitatively assess combinations of NMA and PM data is [12, 13] using a "system-centric" framework applied to conceptual models of an aqueous reprocessing facility. Garcia discretizes all data streams, for example, into "normal," "low," or "high" and currently assumes data streams are independent.

References [14] and [15] report extensive experience with SM data, focusing on monitoring tanks for abnormalities, parsing SM data into key events such as shipments and receipts. See Section VI for more detail. There have been no published attempts to merge SM data with NMA data.

Reference [7] reports on a MatLab/Simulink aqueous reprocessing simulation model that includes measurements of solution flow rates in pipes that leads to a type of "advanced solution monitoring" system. We focus on this type of PM data.

B. *Related Work in Pattern Recognition for Time Series*

There is a tremendous literature on time series and on pattern recognition [16], but relatively little on pattern recognition for multivariate time series [17-19]. An unpublished technical report [19] applied pattern recognition to "unusual" sections of background in multivariate time series.

FIG. 1 A GENERIC AQUEOUS REPROCESSING TANK LAYOUT WITH ONLY A FEW KEY TANKS SHOWN, INCLUDING BUFFER, FEED, RECEIPT, WEIR, INPUT ACCOUNTABILITY TANK (IAT) AND PRODUCT ACCOUNTABILITY TANKS (PAT). THE STAGE NUMBERS INCREASE AS THE PU PURITY AND CONCENTRATION INCREASE FROM SEPARATIONS AND EVAPORATION. LOW ACTIVITY WASTE (LAW) AND HIGH ACTIVITY WASTE (HAW) MUST ALSO BE MONITORED WHEN SHIPPED AND STORED IN STORAGE TANKS

Nma and Pm

A. NMA

The key quantities in NMA are the MB and its measurement error standard deviation σ_{MB}. If the MB at a given time ("balance period") exceeds $k\sigma_{MB}$ with k in the 2-to-3 range, then the NMA system "alarms." Considerable effort is aimed at assessing measurement uncertainties to estimate σ_{MB}. Choosing k in the 2-to-3 range for a low false alarm probability is based on an appeal to a central limit effect arising from combining many measurements to justify assuming the measured MB is approximately Gaussian distributed around the true MB [20-23].

NMA has known limitations, particularly when large amounts of SNM are processed per unit time. Therefore, PM is increasingly important at large facilities [1,2,5,7]. Consider a facility having an input accountability tank (IAT), product accountability tank (PAT), and process operations between the IAT and PAT. If the true PAT output is less than the true IAT input, then the desired safeguards conclusion is "alarm." And, if output is less than input, then various observables must be produced that could be measured. Therefore, PM attempts to verify that material flows and constituents are as declared by looking for the absence of such observables, such as changing material flow rates and constituents to misdirect the SNM to an undeclared exit stream. It is important to understand what types of facility misuse are possible and credible, and also to understand to what extent the various misuse scenarios can be detected.

A sequence of MBs can be evaluated over a fixed period ("period-driven"), or not ("data-driven"), and in either case, the covariance matrix of a sequence of MBs, Σ_{MB}, is estimated. In data-driven evaluation, some type of sequential testing is used, usually including the two basic tests: MUF (material unaccounted for, the same as the MB, which is good for a one-time abrupt loss) and CUMUF (cumulative MUF, which is good for a longer-term loss). Another good choice is Page's test, which is defined at period t as P_t = maximum(P_{t-1} + SITMUF$_t$ – k, 0), where SITMUF is the standardized, independently transformed MUF (should have zero mean, unit variance, and be uncorrelated with all previous SITMUF values), k is a control parameter usually defined to be 0.5 [20-24].

One issue in sequential testing is that the test should have good alarm probability for either abrupt or various types of protracted diversion. The best sequential test depends on the type of loss so no test can be uniformly more powerful for all loss types. The CUMUF test is good if diversion begins on the first balance period and continues at the same rate for all subsequent periods. Page's test is optimal if the diversion begins in an arbitrary period, persists at the same level for an arbitrary period, and then returns to zero. Slight complications arise due to the transformation required (that uses Σ_{MB}) to convert a MUF sequence into a SITMUF sequence [21-24], but Page's test applied to the SITMUF sequence is among the most versatile tests, and is arguably the most versatile [23].

Advantages of NRTA include: (a) improved abrupt loss alarm probability, (b) timeliness, (c) improved alarm/anomaly resolution, and (d) refinement of measurement error models [25, 26, 27]. Regarding measurement error models, metrology for nuclear safeguards includes the notion of random and systematic errors as in the guide to expression of uncertainty in measurement [27,28]. For example, a measured quantity M is assumed to vary around the corresponding true quantity T, with $M = T + R + S$, where R is random error and S is systematic error, and the standard deviation σ_R of R and the standard deviation σ_S of S are estimated using well-characterized standards. Straight-forward variance propagation is then used to estimate Σ_{MB} [20, 22] Regarding SNM in-process inventory that is difficult to measure (called holdup), if there were no measurement error in the transfers and inventory, then the MB would equal the change in holdup plus the true loss. The presence of measurement error complicates MB evaluation, and the presence of nonnegligible holdup together with measurement error further complicates MB evaluation. Nevertheless, provided σ_{MB} is well estimated (not a scientific challenge, but often an engineering challenge constrained by limited time and budget), it is well understood what σ_{MB} implies about loss detection capability.

Remark 1: NMA involves measuring facility inputs,

outputs, and inventory to compute a MB. With a measurement error standard deviation of σ_{MB} = 0.3% of throughput, assuming the measured MB has a Gaussian distribution around the true MB, and international safeguards detection goals (95% detection probability and 5% false alarm probability) the diversion would have to equal 3.3 x 24 kg = 92 kg for an 8000 kg Pu per year facility. This is much larger than one significant quantity (SQ), which is 8 kg for Plutonium.

Remark 2: Facilities that cannot meet the detection probability (DP) goals have negotiated-levels of "additional measures." For example, the Rokkasho reprocessing facility (RRP) in Japan includes PM as a separate, additional safeguards measure.

B. PM

Process monitoring is a broad term that in nuclear safeguards includes monitoring by radiation detectors, cameras, and monitoring solutions in vessels using pressure-sensing dip tube (which is this paper's focus).

NRTA is typically described as: frequent balance closures based mostly on measurements of the shipments and receipts, with varying capability to measure or estimate in-process inventory. In practice, "frequent" is typically daily or weekly (however, PM-based balance closures are common on a per-batch basis which could be daily or multiple times per day). Facilities that close balances very frequently, such as daily or after each batch transfer, rely on various shortcuts or partial measurements. For example, it is rare to equip each processing unit with in-line holdup or in-process inventory monitors. Therefore, either engineering estimates, or historical by-difference estimates are used for negotiated portions of the in-process inventory measurement [29]. In the NRTA scheme at the THORP ("thermal oxide reprocessing plant") in England [23], full material balance closures are not as often as weekly because of the infrequency of Pu concentration measurements. Full balance closures are less often than weekly, but pseudo-balance closures using empirical relations to estimate the Pu concentration are quite frequent (roughly daily). Although in-line dip tubes measure vessel volume every few seconds, there might not be a capability to measure the Pu concentration in-line. In-line dip tubes estimate solution density, so empirical relations together with the density estimate can infer (but not directly measure) the Pu concentration [30] . An NRTA system that measures all material is preferred, but even the best system will typically rely on partial measurements and/or engineering estimates for a least part of the in-process material [10].

Solution monitoring (SM) is a type of PM. Consider level (L), density (D), and temperature (T) measurements of solution in a reprocessing facility. Unless there is an in-line Pu concentration measurement, then empirical relations linking Pu concentration to D and T for a given nitric acid concentration are required to estimate the Pu concentration. Together with a volume estimate using the calibrated V = f (L) + error relation, an estimate of Pu mass is available. This is a pseudo-measurement because unless Pu is actually measured, we cannot be sure that Pu has not been diverted in some manner without reducing solution volumes.

The type of PM just described is essentially a poor-man's NRTA and can lead to high DPs for abrupt diversion. Reference [25] showed that SNM loss during tank "wait modes" would be much easier to detect than SNM loss during "transfer modes" (see Section IV). This is largely due to canceling systematic errors when two level measurements in the same tank are compared. If we need high confidence in PM only during transfer modes, this is a potential savings. However, because there is no in-line Pu concentration measurement, the caveats mentioned earlier in this section are in effect. The adversary could divert without an alarm during a wait mode by replacing the removed volume with the correct density solution. If this occurred over a one day period (the daily Pu throughput is approximately 50 kg), then downstream Pu concentrations could be back at expected values by the next monthly balance closure when Pu concentrations are measured in all key tanks.

To summarize Section III, short-cut assay methods such as a volume and a calculated SNM concentration do not directly measure the SNM of interest but are often used for some of the measurements in frequent NMA (NRTA). PM directly supports NMA if PM is used to estimate holdup [31, 32]. Regarding holdup, if there were no measurement error in the transfers and inventory, then

the expected value of the MB would equal the change in holdup plus the true loss *L*. The presence of measurement error complicates MB evaluation, and the presence of nonnegligible holdup together with measurement error further complicates MB evaluation. Nevertheless, provided σ_{MB} is well estimated, which is often an engineering challenge constrained by limited time and budget, and which often invokes modeling and simulation to estimate holdup and model measurement processes, it is understood [20,22] what σ_{MB} and/or Σ_{MB} implies about loss detection capability.

Event Marking

Raw SM data are unlikely to be useful as input features to pattern recognition. Instead, raw SM data can be parsed into key events such as shipments and receipts, as done by some SM evaluation systems (SMES) [33,34]. This allows us to regard each tank as a sub-MBA (also called a unit process accounting area in [10] and generate residuals that are analogous to the MB from NMA. Alternatively, flow rates to and from tanks can be used to generate very frequent (every few minutes) residuals from each tank, without explicit event marking [7]. To focus this paper, we only consider the event marking option.

Tank-monitoring requires signal estimation and change detection (also called event marking) in noisy scalar-valued time series. Tank data arrives in a streaming fashion, approximately every 1 minute or even more frequently. In-tank temperature (*T*) is measured and in-tank dip tubes at various tank heights measure pressures that can be converted to solution density, level (*L*), and volume (*V*) via a level-to-volume calibration. Mass *M* in the tank is then $V \times$ density. The frequent in-tank measurements can be regarded as (*L*, density, *T*) or (*V*, *M*, *T*). Tank level *L* can be monitored without converting to *V* or *M*. However, during tank-to-tank shipments, solution *V* and *M* are conserved so any scheme to monitor *L* changes in the shipper tank compared to *L* changes in the receiver tank, must consider the level-to-volume calibration. The examples below assume that *V* is the same linear function of *L* in all tanks so it is adequate to monitor *L* changes during tank transfers as a surrogate for *V* changes.

The main goals are to identify and monitor activities in each tank for consistency with historical behavior, and the challenges are sufficiently broad to illustrate several key concepts in signal estimation and change detection. The event-marking approach regards each tank as a material balance area [33], so *V* and *M* changes during transfers are compared to a corresponding upstream shipper tank and downstream receiver tank to monitor for special nuclear material loss. During non-transfers or "wait" modes, one must check for small subtle *V* and/or *M* changes. In practice, anomaly free training data is required to establish alarm limits to monitor *V* and *M* during tank-to-tank transfers and during wait modes [25,34]. It is not anticipated that safeguards personnel would routinely evaluate the large amounts of data generated from monitoring all transfers and wait modes for all tanks. Instead, some type of statistical monitoring system will flag only anomalous transfers and wait modes [33].

At present, the change-detection algorithms are implemented in a somewhat ad-hoc manner in SMESs [33], stepping forward in time checking for "significant" changes while flagging, but otherwise ignoring, known temporary perturbations such as tank sparging and recirculation [33]. In some types of tank sparging, nitrogen is bubbled through the tank to remove oxygen build-up. In other types of tank sparging, air is bubbled in to homogenize tank temperature. In either type of sparging, solution level jumps up, then returns, then jumps up, then returns,…, and sparging leads to increased evaporation. Individual tanks are often vented to a common location (a "header") to which the evaporate travels, and condensate can return to the same tank or to another tank. Tanks can be sparged for approximately 1 minute every 10 minutes. In addition to sparging perturbations and associated evaporation, tanks are recirculated by exporting a significant portion of tank contents to a loop that returns back to the same tank to achieve large scale mixing, often prior to sampling. Recirculation requires pump action that can temporarily increase solution temperature. The type-2 evaluation method described here monitors "wait" and transfer modes for *M* and/or *V* changes. Notice that "wait" is quoted because of the perturbations that occur during "wait" modes.

An important observation is that in real data, tank-to-tank transfers exhibit larger variation than calibration experiments predict [35]. One reason for this large variation is imperfect event marking. Another reason is process variation involving the solution transfer mechanisms that can lead to temporary deposits and withdrawals of solution to and from the pipework connecting the tanks. Although it has been noticed that *V* transfer differences (TDs) between tanks tend to be larger than anticipated on the basis of tank calibration data, prior to this work there has not been an attempt to quantify the effect of imperfect event marking effects on the error contributing to variation in observed volume TDs.

Figure 2 shows realistic simulated true level readings which will be denoted as μ_τ. Figure 3 shows example results of the found and marked events for the data in Figure 2 that is modified slightly by adding simulated Gaussian random measurement errors [8,28]. In Figure 2, these true readings in arbitrary units (au) do not include any measurement error but do illustrate most of the challenges, including the presence of: (1) many changes in rates and changes in durations; (2) different spacings between events such as tanks filling and emptying; (3) nuisance high-noise subevents; (4) break or bend points in true signals that arise due to solution transfer rates changing, and (5) inconsistent event signatures. Events labeled A are typical receipt/ship events. Event B is also a receipt/ship event but the shipment is interrupted before completion and there is evaporation during the "wait" mode. Event C is a tank sparging event. Events D are two sets of erratic measurements due to instrument faults. For example, one instrument fault sometimes arises from the formation of crystals temporarily partially plugging a dip tube. Event E is a recirculation event.

Depending on context, the term "noise" refers to either measurement error or to nuisance changes in true tank level μ_τ such as those that occur in sparging or temporary instrument faults. There will be no attempt to distinguish among such nuisance changes here. Transfer mode involves a shipper and receiver tank. Wait mode involves only one tank, but could involve transient behavior such as recirculation (event E in Figure 1) or evaporation (during the "wait mode" of event B in Figure 1).

Change point literature typically specifies a data model, which is also emphasized in [33]. Figure 2 (top) suggests that, except for some of the nuiscance-change regions, the true levels can be well modeled as piecewise linear or constant. Of course any function observed at discrete time steps is piecewise linear, but the pieces in this application are relatively long time sections reflecting tank activity or inactivity.

Measured readings y_t (which can be regarded as level *L*) will have measurement errors present, and generally there could be both relative and absolute errors so $y_t = \mu_t (1+ S_{Rel} + R_{Rel}) + S_{Abs} + R_{Abs}$, where *S* is systematic error and *R* is random error [35,36]. The bottom figure in Figure 2 plots the relative lag-1 differences $d_t = (y_t - y_{t-1})/y_t = (\mu_t - \mu_{t-1})/\mu_t$ in the case of zero measurement error. Figure 3 shows example results of using d_t on simulated y_t to find and mark events. Only random relative errors were added for the illustration in Figure 3, with a relative random error standard deviation $\sigma_{R,Rel}$ of 0.5%. A custom function find.events.diff in the statistical programming language R [37] is reasonably effective in implementing event marking [33].

The evaporation occurring during the wait mode of the type B event has an exaggerated rate so it is easy to see. Evaporation currently will not be detected as an event, for any anticipated rates of evaporation. However, wait modes can still be monitored for consistency with historical behavior. If small volume loss and very small mass loss typically occur during "wait modes," the anticipated explanation is evaporation.

Events C and D are easily filtered out using a kernel smooth (lokerns in R, see [33]) so can therefore be ignored if desired. Or, if desired, events associated with tank sparging, sampling, recirculation, etc. could be monitored for consistency with historical behavior. To monitor sparging behavior (event C), one can compare raw data to lokerns smoothed data to detect possible sparging regions in order to archive sparging examples to learn historical sparging patterns. Events of type D present a challenge, but we have found that applying find.events.diff to lokerns-smoothed data will nearly always ignore a small event of type D. In addition, the

rates of nuisance-change occurrences such as type D and their typical patterns (how many, how large, and how spaced in time) for each tank can be monitored. Finally, recirculation events, E, are detected using find.events.diff applied to smoothed data. However, recirculation does not involve another tank, and recirculation events can if desired be treated like sparging events by monitoring for agreement with historical patterns.

Figures 3 and 4 illustrate event marking results for relatively simple cases. Figure 3 was described above, and note in Figure 4 that the sample event in the input accountability tank (IAT) can be ignored if desired. For illustration, measurement errors and process variation effects are added in the second portion (but not the first portion) of the example in Figure 4. Even in relatively simple tank-to-tank transfers, the observed volume and mass transfer differences between tanks often exhibit a multi-modal distribution (a common type of non-Gaussian behavior) arising from pump and pipe carryover effects in which some transfers temporarily donate Pu to the pump and pipes and other transfers withdraw Pu from the pump and pipe.

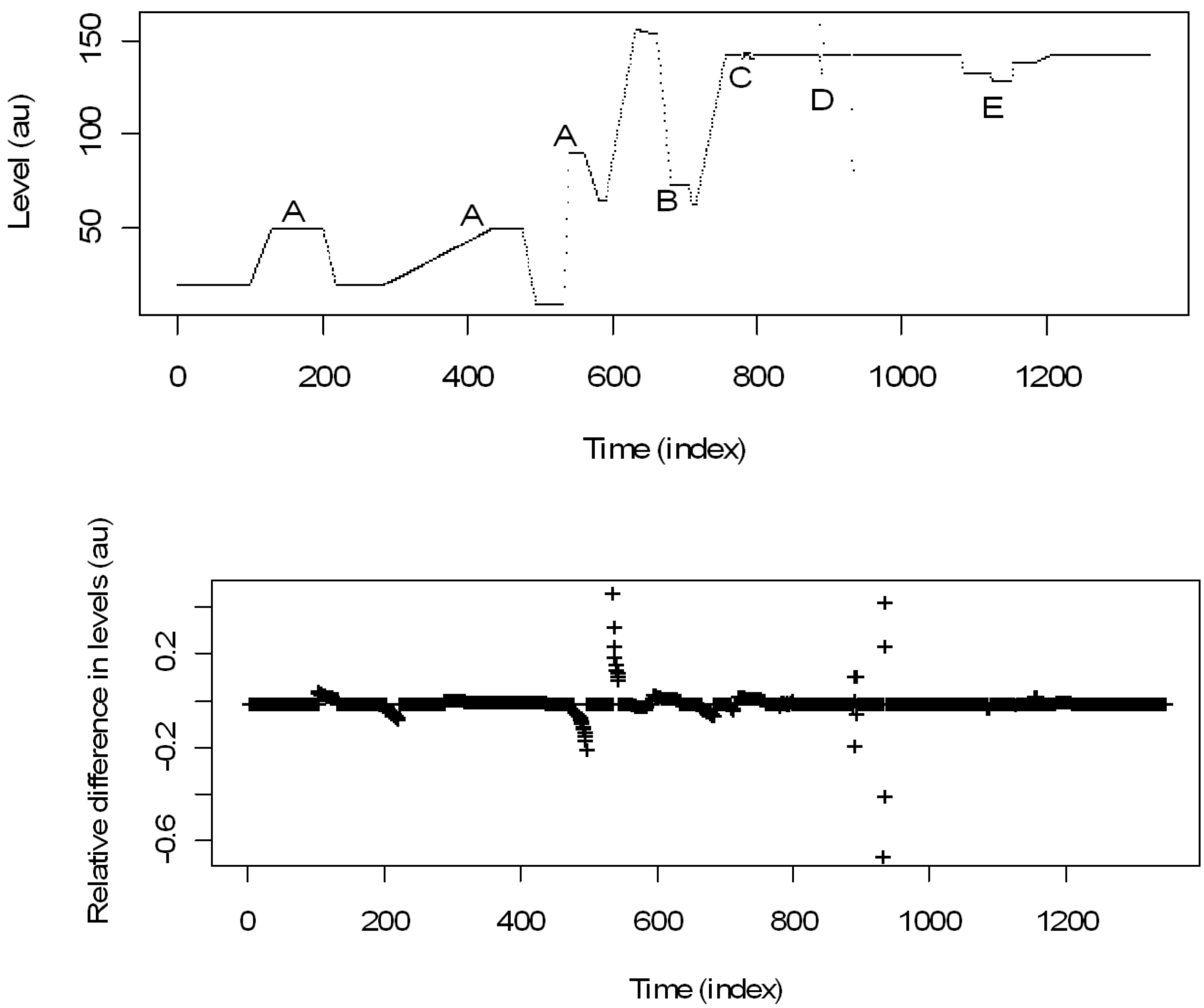


FIG. 2 TANK LEVEL *L* IN ARBITRARY UNITS VERSUS TIME IN (A) AND THE RELATIVE LAG-1 DIFFERENCES IN *L* VERSUS TIME, INDICATING TIMES OF EVENTS STARTS AND STOPS

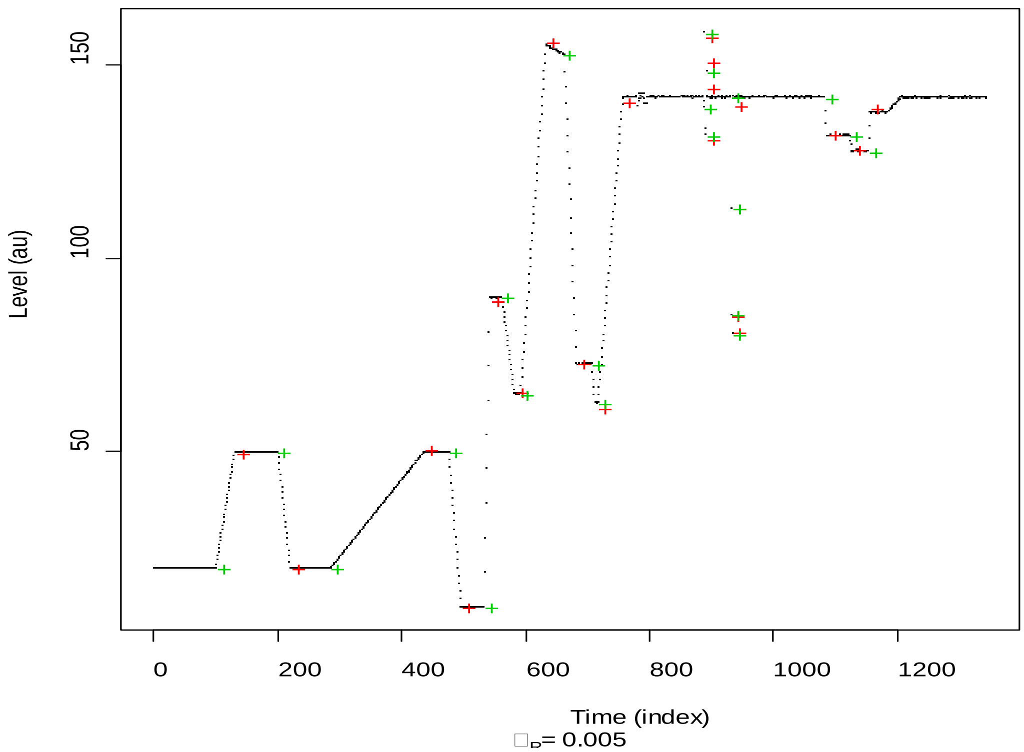


FIG. 3 EXAMPLE OF ESTIMATED START AND STOP TIMES IN EACH FOUND EVENT IN FIGURE 2 ASSUMING A RELATIVE RANDOM ERROR STANDARD DEVIATION σ_R OF 0.5%. THE FIRST (GREEN) "+" IS THE START AND THE SECOND (RED) "+" IS THE STOP OF EACH FOUND EVENT

To summarize Section IV, we aim to ignore recirculation, sampling, etc. and parse raw SM data into "wait" and "transfer" modes. However, if loss occurs during an event such as recirculation, then some signal is generated that will possibly be detected in our analysis (see Section IX) of wait and transfer modes in which we regard each tank as a sub-MBA.

Data Fusion

Currently, NMA is the single objective/quantitative basis for DP assessments, with PM being used in various support roles in support of NMA (see Section III). In NMA, diversion detection probability (DP) is the safeguard's system main figure of merit for a specified diversion amount and time frame. Because σ_{MB} determines the DP (see Section III), via the assumed Gaussian Distribution of the MB, efforts are continually made to reduce σ_{MB}.

In combining PM data with NMA data, we propose to retain diversion DP as the figure of merit, but extend the diversion scenario description from SNM amount and time frame to include how the SNM is diverted. A key task is then to estimate the probability distribution of the combined PM and NMA residuals in the no-diversion case and in the diversion case. The residual probability distribution in the no-diversion case can be estimated by analysis of real facility data, and in the diversion case can be estimated by modeling and simulating the effects of facility misuse on real data. Sections IV and VII-IX give more details regarding the non-Gaussian distribution of PM residuals.

Once the probability distribution is estimated in the no-diversion and diversion cases for the combined NMA and PM residuals, data fusion to combine NMA and PM residuals can be done at the feature, score, or decision levels to reach an overall decision [38]. Here, we perform data fusion at the score level, where the score is the NMA or PM residual.

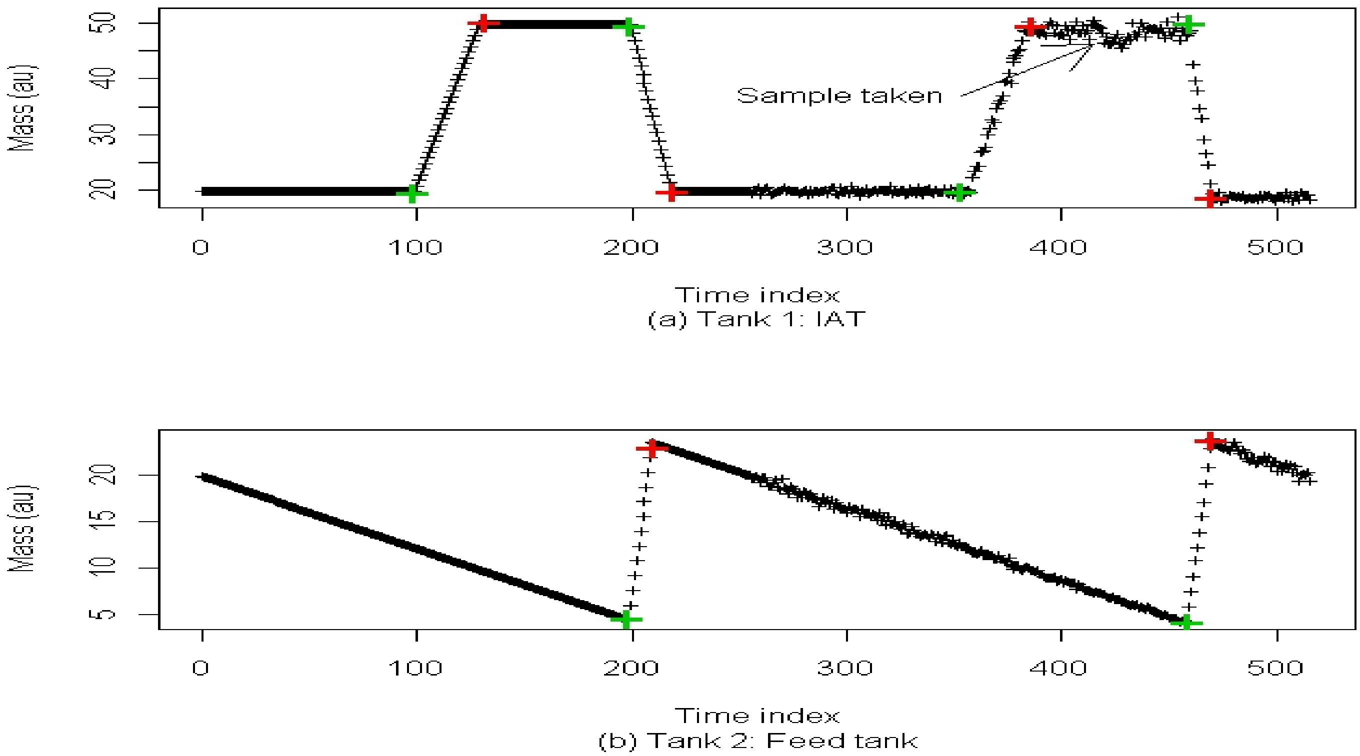


FIG. 4 EVENT MARKING EXAMPLE. THE FIRST (GREEN) "+" IS THE START OF THE EVENT AND THE SECOND (RED) "+" IS THE STOP. THE INTENT IS TO IGNORE SAMPLING EVENTS, BUT TO MONITOR EACH WAIT MODE (WHICH COULD CONTAIN A SAMPLING EVENT) FOR CHANGE. FOR ILLUSTRATION, MEASUREMENT ERRORS AND PROCESS VARIATION EFFECTS ARE ADDED IN THE SECOND PORTION (BUT NOT THE FIRST PORTION) OF THIS EXAMPLE

We propose to estimate a figure of merit for a safeguards system by estimating the system DP from PM combined with NMA using the following steps:

(a) Describe diversion scenarios to inform how data should be evaluated to provide a means of event detection using expert elicitation if possible [39]. Scenarios are characterized by how a specified amount of SNM (in terms of σ_{MB} for ease of comparison to NMA systems alone) is misdirected, and over what time frame;

(b) Extend anomaly resolution work, which has focused on identifying, categorizing, and resolving false alarms [34] to the case of recognizing diversion signatures and examine a variety of pattern recognition/fault detection and diagnosis approaches;

(c) Evaluate P (alarm | diversion scenario), the conditional probability of an alarm for a given scenario. The alarm rule operates on p residuals r_1, r_2, …., r_p which include MB values from NMA, plus residuals from monitoring "wait" and "transfer" modes in tank SM data. The probability P (alarm | diversion scenario) is a function of the true states of nature, the measurement system, and the alarm rule(s). Depending on the desired alarm rule, some subset of r_1, r_2… r_p could perhaps be dichotomized into "exceeds threshold" (1-valued) or "does not exceed threshold" (0-valued).

Each diversion path has signatures (observables), so including relevant PM measurements with NMA data can enable pattern recognition approaches (for example, see the dissolver scenario in Section VIII). We envision two options to combine measurement systems having differing DPs. Option 1 uses a subset (the master) of systems as first alarmers and another subset (the slave) of systems to either resolve the master alarm or to lead to a system alarm. The master system need not alarm if various subsets of the subsystems alarm, depending on

the master alarm rule. If the SM subsystem alarms, then it could also include information such as the time frame over which residuals fed to a sequential test were large, and the tank(s) for which the residuals were large. Such information is useful in deciding whether the master system alarms. If the master system includes NRTA and SM subsystems, then dependencies among the subsystems arise. Both NRTA and SM subsystems use the dip-tube based volume measurements for each tank. Option 2 uses all observables on the same footing, without division into master and slave. Either option could dichotomize the measurements as alarm or no alarm, accept scores from subsystems such as distances from nominal, or accept the raw measurements as input.

SM is an example of very frequent lower quality (higher measurement error and process variation effects) measurements while NRTA involves higher quality less frequent measurements. In a master/slave arrangement, should the NRTA be the master and SM the slave or vice versa? In an "equal footing" arrangement, SM data arrives at a much higher rate, at least several times per day if the event-marking option (Section IV) is used.

For a given scenario, P (alarm at any time 1, 2, …, $t|$ diversion scenario) can be estimated using simulated effects superimposed on real or simulated background data for any SM approach. Lyman [40] points out that not all diversion scenarios can be anticipated and we agree. However, P(alarm at any time 1, 2, …, $t|$ diversion scenario) can be estimated for the scenarios thought to be most credible, and although P(alarm at any time 1, 2, …, $t|$ diversion scenario) cannot be estimated for unspecified scenarios, statistical tests (see Sections VI,VII, IX) can be used that detect any shift in a probability distribution, so we can safely claim that at least P(alarm at any time 1, 2, …, $t|$ diversion scenario) is not zero against any credible but unspecified scenario.

This section has given a broad framework, but to focus here, very specific residuals r_1, r_2… r_P will be used from NMA and PM in the two examples in Section VIII using option 2. In fusing NMA and PM data, recall that NMA uses Page's sequential test to detect trends over time [23, 24]. In Section IX we also use Page's test to define residuals that can detect trends over multiple wait and/or transfer modes for a given tank or pair of tanks.

Hybrid Of Period-Driven And Data-Driven Pattern Recognition

A. *Period Driven Hypothesis Testing*

Suppose NMA and PM residuals are evaluated frequently (every 10 days for NMA and as-generated by event marking for PM), but a statistical decision is made every year to alarm or not. Yearly decisions are popular and practical in safeguards because facilities often schedule a partial shutdown and clean out of the facility, which provides a convenient time to have most SNM in relatively easy-to-measure forms.

One goal for international safeguards using period-driven testing with a one-year decision period is to detect a loss of a significant quantity (SQ) with probability 0.95 with a 0.05 false alarm probability per year, testing for loss only, not for gain, so one-sided hypothesis testing is used. Assuming the MB is approximately Gaussian distributed, one can achieve a DP of 0.95 to detect a diversion of 3.3 σ_{MB} using period-driven NMA with yearly balance closure (non-sequential testing), where the alarm threshold is 1.65 σ_{MB}. However, suppose the adversary diverts material over months 7 to 18, straddling two balance periods (year one and year two). For the system to fail, the system must fail to detect the diversion of 1.65 σ_{MB} in year one, and fail to detect the diversion of 1.65 σ_{MB} in year two, which occurs with probability $\frac{1}{2}\times\frac{1}{2}=\frac{1}{4}$, so the DP is reduced from 0.95 to 1-0.25 = 0.75 [41]. Therefore, the adversary can reduce the DP from 0.95 to 0.75 simply by diverting SQ/2 during year one and SQ/2 during year two.

B. *Data-Driven Hypothesis Testing*

To mitigate a decrease (from 0.95 to 0.75 in the Section VI.1) in DP arising from the adversary diverting across two balance periods, from month 7 to month 18, one can instead use a sequential (data-driven) test that has no fixed period at which decisions are made. Instead, the test continues until a decision to alarm or not is made, and then starts over. We can design a sequential test to have a long average run length (ARL) between false alarms, such as 20 years, which corresponds to the 0.05 per-year FAP assumed in the previous paragraph. One well-known and effective sequential test is Page's cusum test defined at period t as $P_t = \text{maximum}(0, P_{t-1} + y_t - k)$,

where y_t is the SITMUF sequence and k is a user-chosen control parameter that is optimal for detecting a shift from mean 0 to mean $2k$ at an arbitrary period. Page's cusum applied to an independent and identically distributed sequence of N(0,1) random variables (such as the SITMUF sequence) has a DP of approximately 0.79 for this total loss of $3.3\sigma_{MB}$ spread evenly over months 7 to 18 (across balance periods 1 and 2 in period-driven testing) if the ARL is approximately 20 years and $k = 0.5$. And, if the loss is on any one balance period, the DP using Page's cusum is approximately 0.99 (on the basis of 10^4 simulations in R, ensuring approximately a 20-year ARL between false alarms), but is 0.95 for the period-driven yearly balance. If there is a total loss of $1.65\sigma_{MB}$ on a single balance period during year one, then the period-driven yearly balance DP is only 0.50, while the DP using Page's cusum is approximately 0.96, again with a 20-year ARL. There is no avoiding the fact that protracted diversion has lower DP than abrupt diversion, but Page's test manages to retain high DP for abrupt loss while having reasonable DP for protracted loss.

C. *Hybrid of period-driven and data-driven testing*

In Section IX an example with PM and NMA residuals is given in which a period-driven decision is made every 30 days to alarm or not. As shown in Section VI.1, such period-driven testing does not have good DP if the adversary diverts modest amounts of SNM over multiple decision periods. Therefore, if period-driven testing is used, we advocate, in addition, data-driven monitoring of a scalar or vector-valued residual from each period. A scalar residual could be monitored over multiple 30-day periods using Page's cusum as described. A multivariate residual can be monitored using a multivariate sequential test, such as Crosier's cusum, which is a multivariate version of Page's cusum [42].

To summarize Section VI, we propose using a combination of period-driven and data-driven hypothesis testing.

Pattern Recognition

In a typical pattern recognition problem, the data consist of n cases of (y, X) pairs where the integer $y \in (1, 2... J)$ is the class and X is a p-dimensional predictor vector. The goal is to use X to predict the class y, and this task is sometimes called classification, discriminant analysis (DA), or supervised learning. Regarding notation, vectors and scalars can be distinguished by context and definition. For example, y is a scalar and X is a p-dimensional vector.

There are many approaches to pattern recognition. Some attempt to estimate the probability density of the predictor vector, X, given its class (i.e., the class conditional probability, $P(X|y)$) by assuming some convenient distribution for $X|y$ such as multivariate Gaussian which linear discriminant analysis (LDA) assumes [43, 44, 45]. Other methods of estimating densities assume only that the distribution is stationary over time. Such methods are typically called non-parametric or distribution-free methods [46]. Space does not permit a review of all pattern recognition options.

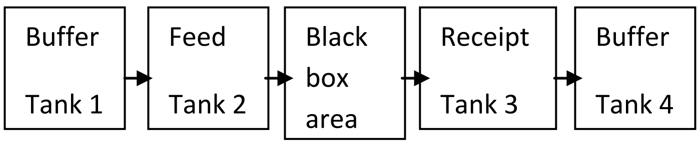
Probability density estimation was invented during the 1950s in order to apply non-parametric DA techniques. Most efforts have focused on the case in which all the predictors are real-valued (continuous predictors). Reference [46] provides an exception that handles real-valued and categorical (unordered and ordered) predictors via density estimation. And, Bayes networks [44] are another option to perform density estimation, but assuming that some of the conditional probabilities in the joint density $P(X|y)$ of X conditional on the class y are known or can be estimated.

Alternative strategies attempt to estimate Bayes rule without estimating the class conditional probabilities, such as support vector machines (SVMs), which construct nonlinear decision boundaries for the classes in a manner similar to flexible discriminant analysis (FDA). Hastie et al. [43] describe SVMs, FDA, and also describe nearest neighbor classifiers and learning vector quantization.

The most common pattern recognition data model assumes that a categorical response y depends on a fixed-dimension predictor X. The pattern recognition task is to estimate $f(X) = \text{Prob}(y = 1 | X)$. The most well studied version of this task assumes the following: (1) all components of X are real-valued; (2) X has fixed dimension, and (3) training cases consisting of (X, y) pairs are independent.

A. *Pattern Recognition for NMA and PM Data*

Currently, most safeguards conclusions are made at the end of a NMA balance period, but the increasing role of PM is driving a change to make data-driven conclusions. As an example, consider a 4-tank balance area consisting of a buffer tank 1 which ships in batch mode to a feed tank 2 which continuously feeds a "black box" area where chemical processing occurs. The black box ships continuously to a receipt tank 3, which ships in batch mode to a buffer tank 4 as in the following diagram with arrows indicating material flow direction.



For process control reasons, the plant operator periodically samples tanks 1 and 4 to measure the SNM concentration and uses mixing rules and measured flow rates to estimate the Pu concentration and mass in tanks 2 and 3. Online measurements of tank level (which is calibrated to volume), density, and temperature are available every few seconds, so tank volume V and mass M (mass = volume × density) are available every few seconds from each tank. These (V, M) measurements are PM measurements. NMA computes the MB as estimated Pu into tank 1 minus the estimated Pu out of tank 3. There are also neutron detectors in the black box area to monitor Pu inventory in an indirect semi-quantitative manner.

The pattern recognition tasks are: (1) to recognize any departure from normal process operations, and (2) to recognize specific misuse scenarios that are judged to be credible. Some of the technical challenges are:

- For (1), anomaly detection as a special case of pattern recognition has been approached using density estimation [45];
- For (2), signatures and patterns of specific misuse scenarios are usually modeled and there is consider model uncertainty, so the probability density function (pdf) of each misuse scenario is uncertain (this source of uncertainty is currently ignored);
- PM measurements overlap with NMA measurements (example: the same instrument that measures tank V for NMA is used for PM) so there are between-data-type correlations;
- PM and NMA data are on differing time scales, and
- PM data captures many innocent sources of process variation.

The main task for pattern recognition is to combine residuals from NMA and PM to provide data-driven pattern recognition (operating as declared or misuse A or misuse B), period-driven (at the end of each day or balance period, make a statistical decision to alarm or not) pattern recognition, and some type of hybrid of period-driven and data-driven pattern recognition as discussed in Section VI.

Remark 3. All predictors for pattern recognition will be based on model fitting and associated residuals. As in "phase 1" control charting [47,48] for production processes, the probability density function (pdf) of the time series of a vector of residuals can be estimated. However, estimation of the residual vector's pdf requires a combination of modeling and data analysis as illustrated by example in Section IX [4]. Our approach described in Section VI and illustrated in Section IX does not distinguish sensor faults from SNM loss, but assuming no more than one sensor malfunctions within small time windows, Howell et al. [15] and Hines et al. [49] illustrate options that are also based on monitoring residuals, using regression and other statistical tools that were first applied to monitor sensor health for the Nuclear Regulatory Commission. To the best of our knowledge, only Howell et al. [15] have attempted to distinguish sensor faults from SNM loss.

Recall that in NMA alone, the figure of merit is P(alarm|L, time period) where L is the SNM loss (due to diversion or innocent loss). And, the central limit effect operating on the many measurements comprising a MB leads to the MB having approximately a Gaussian distribution, so P(alarm|L, time period) for a given alarm threshold is a function only of σ_{MB}. In period driven testing, the time period is fixed in advance, such as one year, and [9] showed that in the Gaussian case, a single CUMUF test at the end of each time period has the highest DP for the worst-case diversion. And, the worst-case diversion vector L is proportional to the row sums of Σ_{MB}. In data-driven (sequential) testing, the time period must be specified for each diversion of interest in order to estimate P(alarm|L, time period), and more complicated alarm rules than the CUMUF rule must be used, such as Page's cusum. Both in model-based

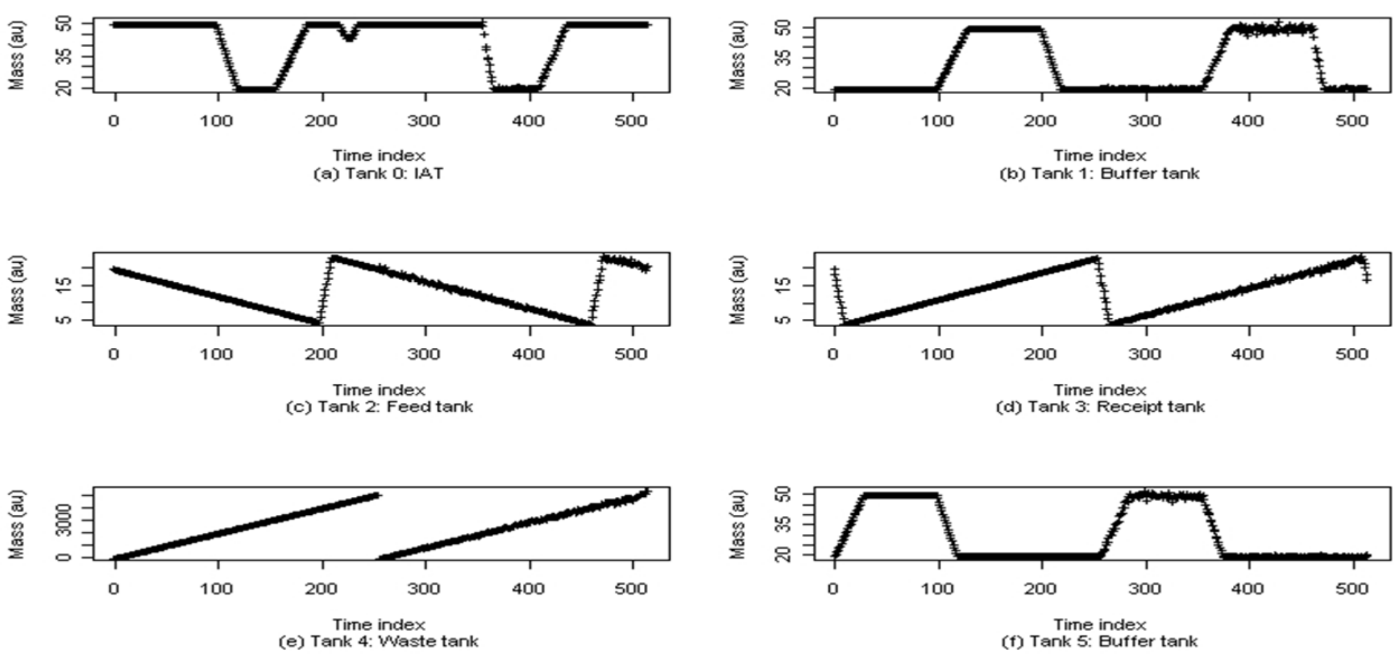
prediction (Section VIII) and in event-marking-based PM (Section IV), the PM residuals will not be adequately modeled using a Gaussian distribution, which complicates the required pattern recognition task. In addition, with time series of combined PM and NMA residuals, either hybrid or pure data-driven testing will be used in the context of evaluating P(alarm|L, time period), where how the diversion occurs must also be specified.

To summarize Section VII, unique aspects of pattern recognition in the context of diversion detection in multivariate time series of PM and NMA residuals were briefly described.

Model-Based Prediction For The Snm In Effluent Streams

SM (perhaps extended beyond in-tank level, density, and temperature to include flow measurements and/or in-line Pu concentration measurements) can help provide a predicted or book value for waste streams. For example, Bakel et al. [5] describe a model for the head end of an aqueous reprocessing plant that results in a model-based prediction (or "book value") of the Pu mass in the hulls waste stream. Xerri et al. [50] distinguish holdup from "hidden inventory" and use by-difference PM data to estimate holdup. Assuming that diversion of excess Pu to the hulls is the only credible diversion route in the head end, it is valuable to have such a "model-based" prediction of the Pu in each hull batch that relies on easily measured quantities such as dissolver cycle time, temperature, and feed nitric acid concentration or bulk density. Similarly, pulsed column models [11,31,32] can provide a book value for effluent streams (an example is given in Section IX). The intent is to detect off-normal conditions that could indicate misdirection of Pu. Monitoring such profiles can lead to residuals as we have described for simpler models involving mass and/or volume balancing of SM data for each key process tank.

Model-based predictions as just described can provide a new way for PM plus NMA to detect diversion on the basis of monitoring the corresponding residuals. A key fact is that diversion to streams that should have relatively small amounts of Pu can be easily detected provided frequent PM data is available, and the model-based predictions are reasonably high quality (i.e., have low total error variance).



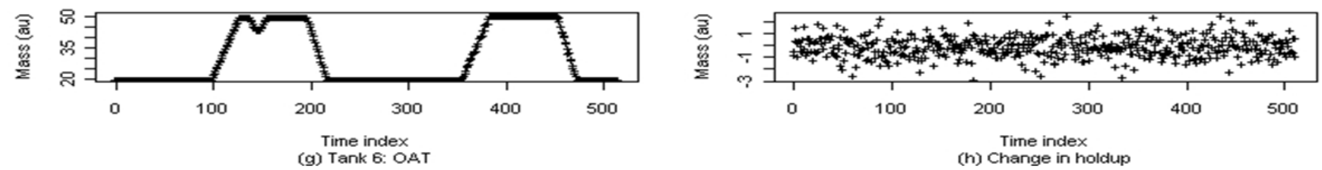


FIG. 5 TANKS 0-6 IN A 7-TANK MBA IN (A)-(G) AND THE MEASURED CHANGED IN HOLDUP IN (H). PROCESS VARIATION AND NOISE IS ADDED IN THE SECOND PORTION OF THE TIME SERIES. ALL ANALYSES REPORTED USE THE NOISY DATA. TANKS 1, 2, 6, AND 7 ARE ALL BATCH RECEIPT AND SHIP (B/B MODE). TANKS 3, 4, AND 5 EACH HAVE A CONTINUOUS (C) MODEL

Examples

We consider a 2-tank and then a 7-tank material balance area. Figure 5 is simulated tank volume *V* data versus time for 7 tanks.

2-tank Material Balance Area

Figure 6 displays scaled volume residuals versus time for wait and transfer modes for tanks 1 and 2 only, in a 2-tank MBA. An MB is computed every 10 days, and 30 days of operation is shown, so 3 simulated MBs are also scaled and plotted. Both the volume residuals and the MBs are scaled by dividing by the respective standard deviations (V/σ_V and MB/σ_{MB}, respectively).

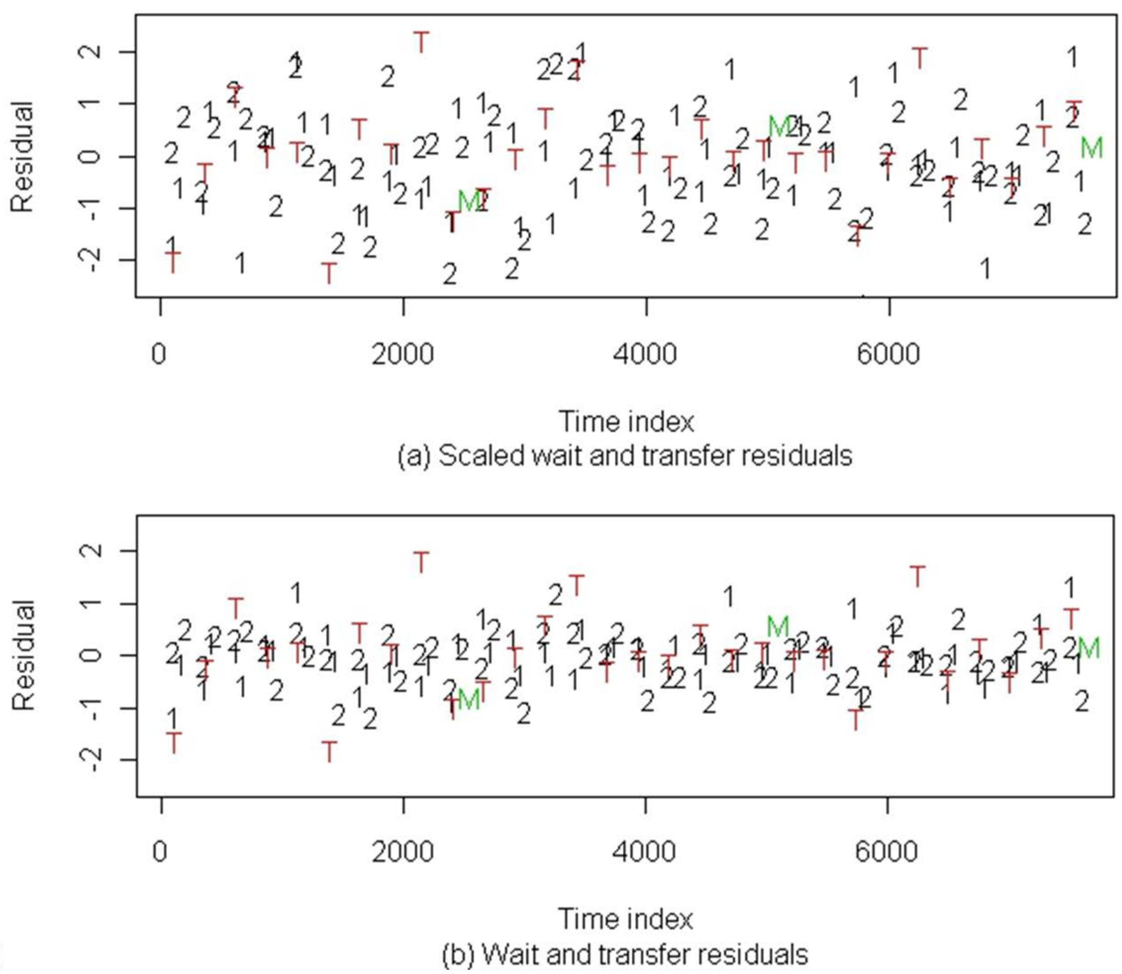


FIG. 6 RESIDUALS FROM MONITORING "WAIT" AND "TRANSFER" MODES IN TANKS 1 AND 2, AND THE MATERIAL BALANCE ("M") AT DAYS 10, 20, AND 30. EACH INCREMENT OF THE TIME INDEX IS 6 MINUTES. INTEGERS 1 AND 2 ARE WAIT MODE RESIDUALS FOR TANKS 1 AND 2, RESPECTIVELY. "T" IS THE TANK 1 TO TANK 2 TRANSFER RESIDUAL (SHIPPER-RECEIVER DIFFERENCE), AND "M" IS THE MATERIAL BALANCE AT DAYS 10, 20, AND 30

In Figure 7, the DPs for each of three commonly-used statistical tests for diversion are plotted versus the mass of Pu lost in units of σ_{MB}, which are the diagonal entries in the 3-by-3 Σ_{MB} matrix (for the 3 10-day balance periods). The three tests are Page (Section VI), cusum (Section VI, the sum of the three MBs over the 30-day period), and Shewhart or "Max" test (if the maximum of the three MBs exceeds a threshold, then alarm). Figure 8 is the same as Figure 7, but for illustration the sign was reversed to negative for the lag-1 off-diagonal entries and other off-diagonal entries were set to 0. Such a tri-diagonal Σ_{MB} is sometimes evaluated in safeguards studies involving relatively large inventory compared to per-period throughput, resulting in a classic tri-diagonal form with negative off-diagonals [21]. Notice that DPs are higher for this particular tri-diagonal Σ_{MB} than for the corresponding diagonal Σ_{MB}, which is opposite to the pattern in the Figure 7 DPs.

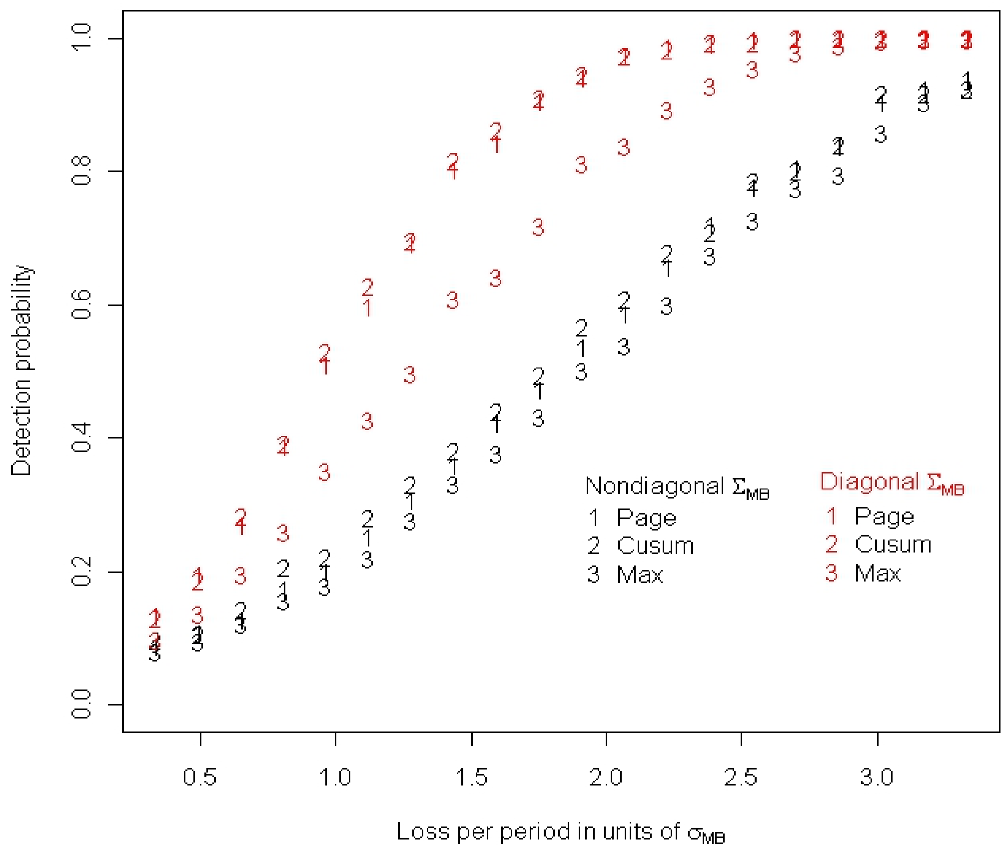


FIG. 7 DETECTION PROBABILITY DP VERSUS LOSS PER BALANCE PERIOD (IN UNITS OF σ_{MB}) ASSUMING INDEPENDENT MBS AND NOT INDEPENDENT MBS (WITH POSITIVE OFF-DIAGONAL ENTRIES) AS DESCRIBED IN THE LEGEND. THE ENTRIES IN Σ_{MB} ARE ESTIMATED USING 1000 REALIZATIONS OF THE RANDOM AND SYSTEMATIC ERROR MODEL.

Figure 9 compares DPs using the Page, Cusum, and maximum alarm rules for various sized losses for short (3), medium (30), and long (100) balance periods. Note that Page's test has the second highest DP, which is often the case with Page, and which is why Page's test is a good compromise choice. That is, the DP with Page's test is often reasonably close to the highest possible DP for a wide variety of loss scenarios. The maximum (Shewhart) test can be applied to the MB or SITMUF sequence. Here the maximum test was applied to the MB sequence.

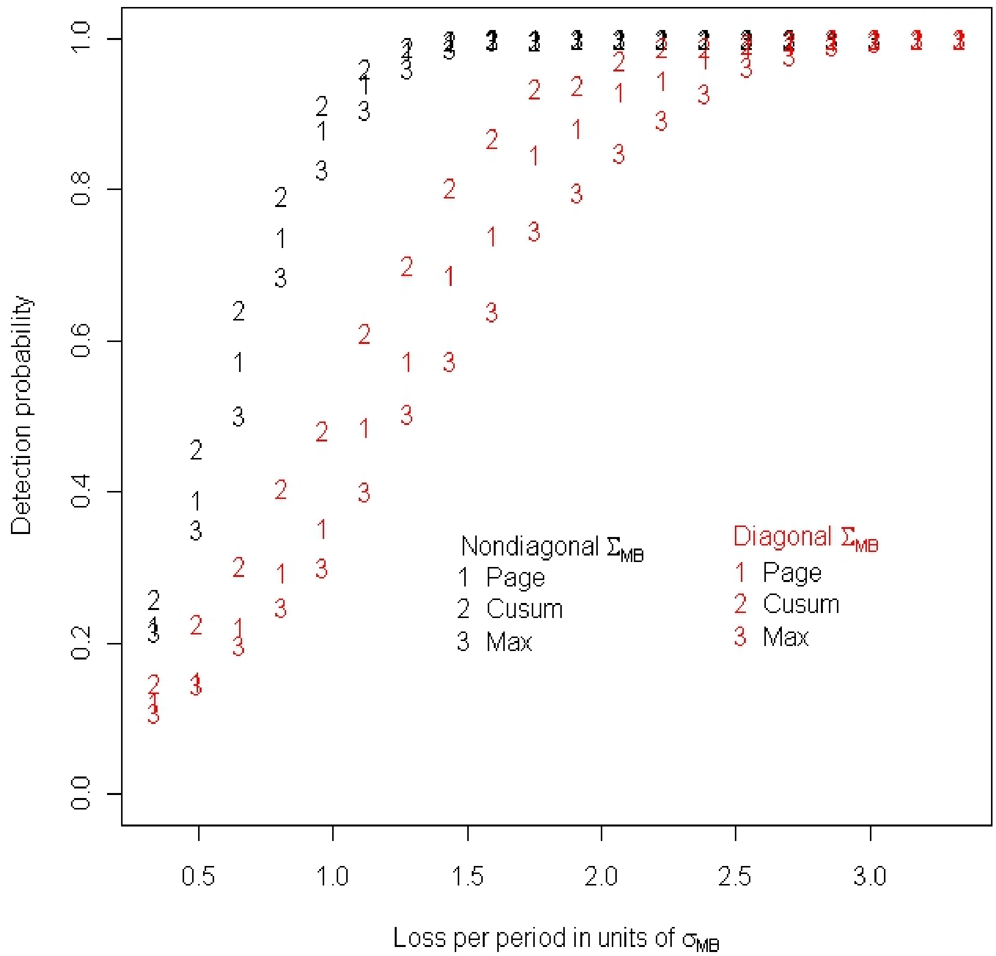


FIG. 8 SAME AS FIGURE 7, BUT WITH NEGATIVE OFF-DIAGONAL ENTRIES IN Σ_{MB}.

Remark 4. Page's sequential test has close to the highest possible DP for many loss scenarios. Any test will have the best DP for at least one loss scenario, which partly explains why so many sequential tests have been proposed for NMA [21].

Remark 5. The Cusum test $C(t)=\sum_{i=1}^{t} MB(i)$ which sums all MBs since the last period ignores individual transfers from tank 1 to tank 2 and has the highest DP among all possible tests for this equal-loss-per-balance-period example [9]. This means that evaluating each tank-to-tank transfer has lower DP than comparing the sum of tank 1 transfers to the sum of tank 2 transfers. Analogously, there is no free lunch regarding the use of SM and NMA data. That is, including SM data is an extension of NMA to include more sub-MBAs (each tank is a sub-MB area) and more frequent balance closures. Therefore, there are scenarios for which using NMA data alone leads to the highest DP. Such scenarios will involve widespread diversion over multiple tanks and time periods (unless such scenarios produce observables that could be monitored, which we are not considering here). The motives for evaluating SM data include resolving NMA alarms (Section III.1), detecting diversion to waste streams that should have relatively small amounts of Pu, and improving abrupt loss detection over more scenarios, meaning that there can be at least moderate DP for a wide range of diversion scenarios, which is not true for NMA data alone.

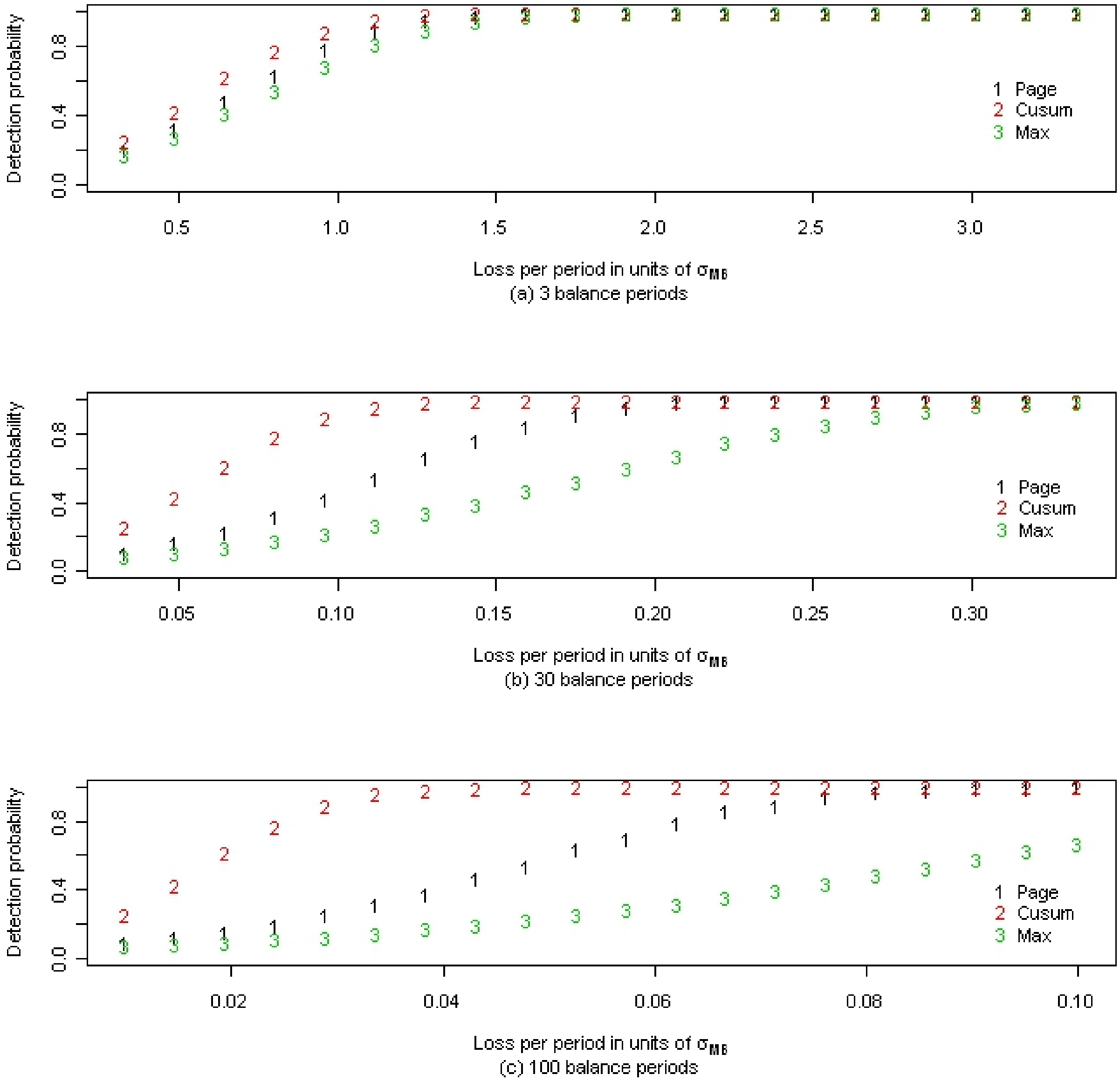


FIG. 9 DETECTION PROBABILITY VERSUS THE AMOUNT OF SNM LOSS PER BALANCE PERIOD ILLUSTRATING THE EFFECT OF LONGER BALANCE PERIODS ON DETECTION PROBABILITIES

A. 7-tank Material Balance Area

Here we extend the 2-tank MBA to 7-tank MBA. Recall that Figure 5 plots simulated data with measurement error and some process variation [35] for the 7 tanks.

This 7-tank MBA includes batch receipt and batch ship tanks (B/B mode) plus batch receipt and continuous ship tanks (B/C) tanks, plus continuous receipt and batch ship tanks (C/B) and holdup and waste. Both holdup and waste have “book values” provided by a model of the pulsed column operation, which is in the separations area between tanks 2 (Feed) and 3 (Receipt). The notion of a predicted value for the waste stream exiting the separations area and for holdup in the separations area

leads to additional residuals to monitor (see Section VIII). These residuals will not be independent of the residuals from NMA (MB values) or of some of the residuals from wait and transfer mode monitoring in SM.

Remark 6. Pu mass measurements in waste streams are a component of the MB, and these same measurements of waste stream Pu mass can be compared to the model-based prediction (see Section VIII), resulting in two correlated residuals, one score being the MB and another score being the comparison between book and measured waste stream Pu mass. Efforts are underway to resurrect and improve models of the separations cycles, but for our purposes here, the separations cycle model to estimate Pu mass is assumed to provide a total relative error standard deviation of 10%. This estimate can be compared to the by-difference estimate of Pu mass in the separation cycle that is obtained by monitoring the SNM flows in and out of the separations area (using tanks 2 and 3) [11,31,32]. Recall that additional residuals arise in the approach where each tank is a sub-MBA and is monitored for M and/or V loss during all "wait" and "transfer" modes.

Figure 10 displays residuals from monitoring each tank's wait modes and all tank-to-tank transfers, from the three MBs over 30 days (one MB every 10 days), and from comparing three SM-based measurement to each of the three "book" values for holdup and for waste. One main challenge is to combine correlated multivariate NMA and SM residuals such as shown in Figure 10 into an overall system having small false alarm probability and reasonable large DP for a range of diversion scenarios. Figure 11 illustrates zero and non-zero correlations between 200 simulated realizations of one pair of residuals in Figure 10. The measured transfers from the IAT (tank 0) to tank 1 are correlated with each MB. This is not surprising, because the IAT measurement error makes a significant contribution to the MB.

An example of how one might combine NM and SM data is plotted in Figure 12 using principal coordinates (PC) [43] to display scores from 19 separate Page's test values applied to 19 residuals from NMA and SM over 30 days spanning 3 10-day NMA balance periods. The 19 residuals include 8 wait and 2 transfer mode residuals, 3 waste measurements compared to the waste predicted value, 3 holdup estimates based on SM data compared to the corresponding holdup measurement, and 3 MBs. The 8 wait mode residuals arose from treating the "high-tank-level" wait modes and the "low-tank-level" wait modes as separate residuals. This resulted in 8 wait modes because 4 tanks each have 2 wait modes. Also, for this 19-score option, only 2 the batch ship/batch receive mode tanks were monitored (avoiding the more challenging transfer modes associated with continuous-mode tanks), in transfer mode and 3 residuals from both holdup and waste and 3 MBs.

Qualitatively, we see that the combined NM and SM data has moderate DP for the moderate loss and large DP for the large loss (the moderate and large losses are shown in Figure 15). The Mahalanobis distance from the zero-mean (zero loss) case could be applied as a simple pattern recognition method (equilavent to DA) to quantify the DP [4, 16, 43, 51].

Both period-driven and data-driven options have been evaluated. For period-driven, 10-day balance periods are used, to illustrate, without attempting to optimize balance closure timings, which in this example resulted in 33 scores (see below for a description of these 33 scores, where "score" is a slight generalization of "residual," because it can include the value of Page's cusum applied to the residual time series). For a data-driven option, Page's cusum is applied to individual data streams regardless of the balance-period timing, resulting in 19 scores as described above.

Figures 13 and 14 are similar to Figure 12, but are both for a widely distributed loss across wait and transfer modes from many tanks over many batches. Notice (compared Figure 13 (a) to Figure 14 (a)) that using the MB sequence alone is most effective for this widely distributed loss [9]. The residuals used to compute the value of coordinates 1 and 2 Figures 12-14 are simulated assuming zero loss and a large loss during each of 5 wait modes for tank 7 (PAT).

In this example, we did not distinguish between bulk volume or mass residuals and Pu mass residuals, and for simplicity here, one can assume that all residuals are Pu mass residuals. Alternatively, if some residuals are bulk volume or mass residuals, the analysis steps remain the same (but the system is more vulnerable to diversion-with-solution-replacement scenarios in which bulk properties are maintained while Pu is removed).

Appendix 1 provides flowchart of the 5 analysis steps in this approach to the 7-tank MBA example.

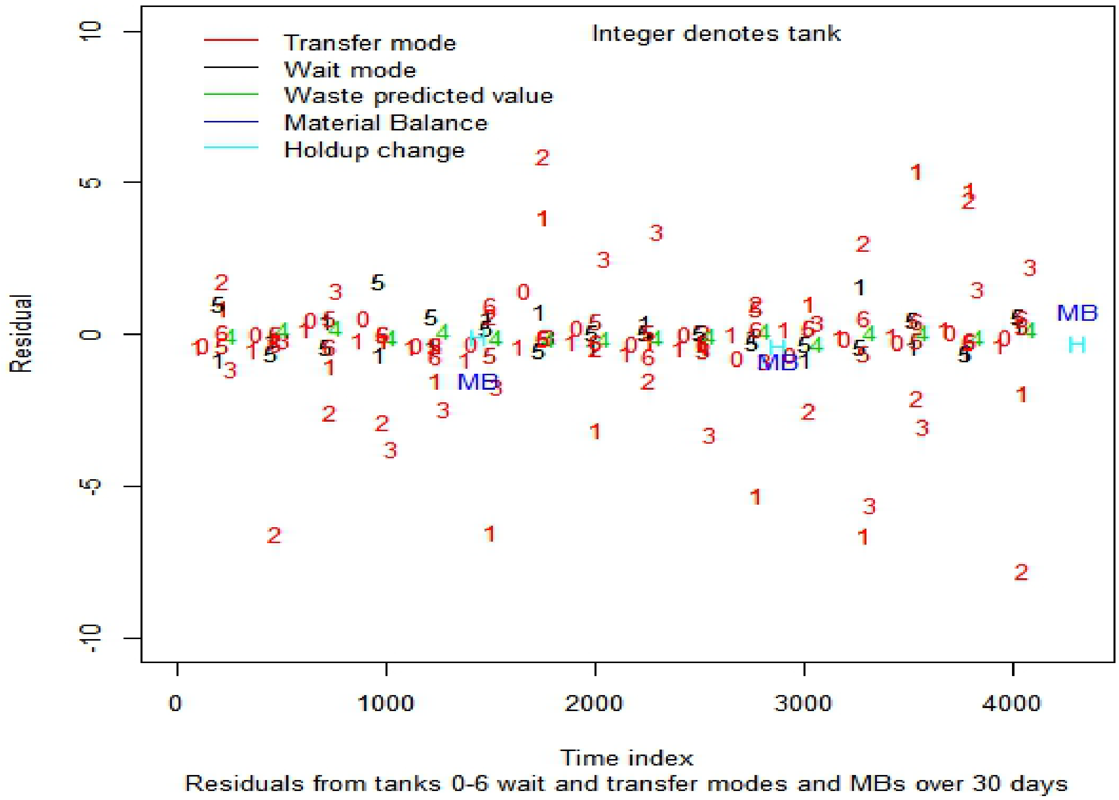


FIG. 10 RESIDUALS VERSUS TIME OVER 30 DAYS. ALL RESIDUALS RESULT FROM A TYPE OF MASS BALANCE. THERE IS A "BOOK VALUE" FOR EACH WASTE BATCH AND A "BY-DIFFERENCE" HOLDUP-CHANGE MEASUREMENT COMPARED TO A NEUTRON-ASSAY HOLDUP-CHANGE MEASUREMENT. FIGURE 10 IS SIMILAR TO FIGURE 6, BUT INCLUDES MORE TANKS AND SCORES FROM WASTE AND HOLDUP MONITORING

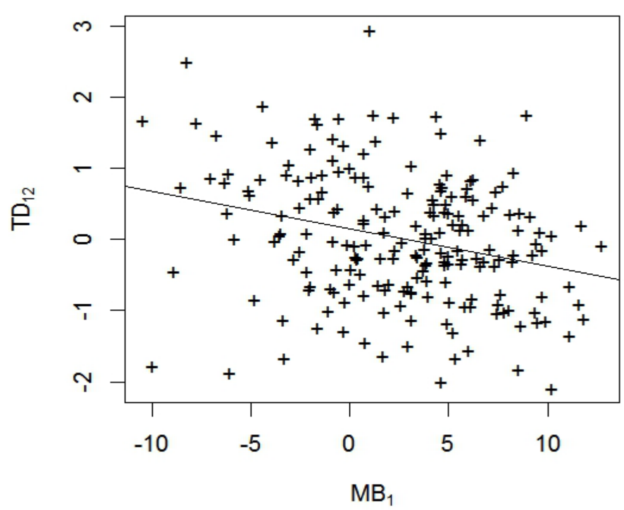


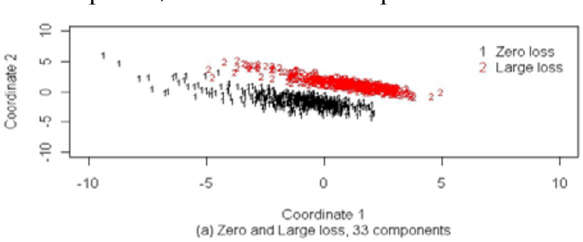
FIG. 11 EXAMPLE OF NON-ZERO CORRELATION (APPROXIMATELY -0.3) BETWEEN PAIRS OF 200 SIMULATED REALIZATIONS OF THE 30 DAYS OF RESIDUALS FOR THE 7-TANK MBA SUCH AS SHOWN IN FIGURE 10. THE TRANSFER DIFFERENCE TD_{12} IS BETWEEN TANKS 1 AND 2 AND MB_1 IS THE FIRST MB, ON DAY 10

Figure 15 plots example DPs for a moderate and large loss for two monitoring options for the residuals in Figure 10. Option 1 uses 13 separate Page tests, one for each of the 10 wait and transfer modes, and one for each of the waste, holdup, and MB sequence. Option 2 uses the Mahalanobis distance from mean of zero-loss distribution. The DP results for the small (near 0.01 to 0.05) FAPs are most relevant. DPs for the higher FAP are given for completeness and to emphasize that it is important to control for the FAP in "multiple-testing" situations.

To illustrate Page's test, Figure 16 plots the Page statistic for some of the residuals such as in Figure 10 over the 30 days. Notice that the tank 7 transfers (out of the MB) are alarming because the loss was simulated from tank 7.

Table 1 lists DPs for small, moderate, and large losses (from tank 7 wait mode only, or widely distributed, as in Figures 12-14) for the 19-score example (expanded to 33 "scores" by including the value of Page's test at all three balance periods for all residual streams, see the next paragraph) using the Mahalanobis distance. For the distributed loss over all tanks, note that the APs are low, even for the larger loss. However, if only the 3 MBs are used (rather than all 33 scores), then the DPs are much larger, as shown in the last 3 rows of Table 1. Also, note that for the large loss from tank 7 wait mode only, DPs are much higher using the 33 scores with the Mahalanobis distance. Here again we have used the term "score" as a slight generalization of "residual," because in some cases, the monitored quantity is the value of Page's statistic at a given balance period.

The DP results in Table 1 are for the 33 scores expanded from the 19 scores that arise from Page's test applied to each of 4 wait modes (one for each of tanks 1, 2, 6, and 7), 4 transfer modes (tank 1 to tank 2, tank 2 to tank 3, tank 4 to tank 6 and tank 6 to tank 7), waste score, holdup score, and MB for each of 3 balance periods. Page's test applied to 4 wait mode residuals and to the 4 transfer modes residuals results in 12 + 12 = 24 scores over the 30 days, and the other 9 scores (24 + 9 = 33) are from Page's test applied over 3 balance periods to the waste stream, the holdup area, and to the MB sequence.



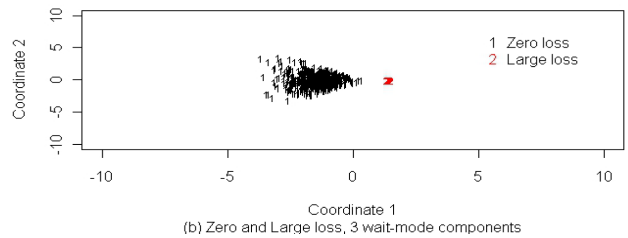


FIG 12 QUALITATIVE ASSESSMENT OF THE ABILITY TO DETECT A LARGE LOSS USING (A) ALL 33 COMPONENTS, OR (B) USING ONLY 3 WAIT-MODE COMPONENTS FROM THE PAT

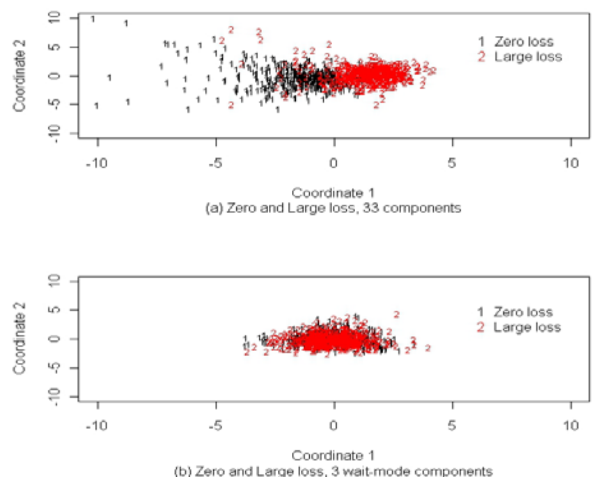


FIG. 13 SAME AS FIGURE 12, BUT FOR A WIDELY DISTRIBUTED LOSS ACROSS MULTIPLE WAIT AND TRANSFER MODES FROM ALL TANKS

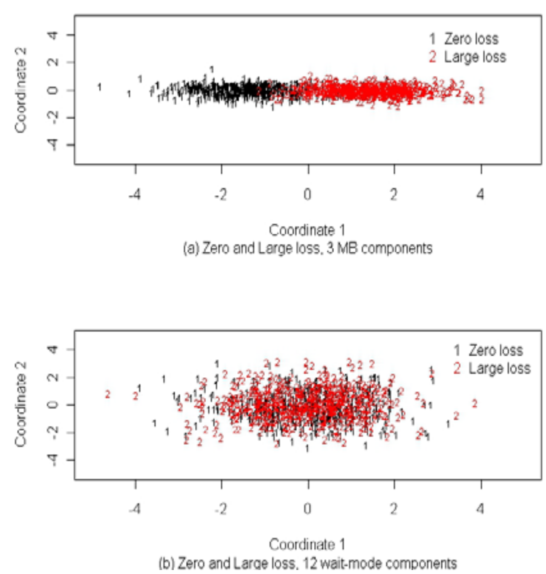


FIG. 14 SAME AS FIGURE 12, FOR A WIDELY DISTRIBUTED LOSS ACROSS MULTIPLE WAIT AND TRANSFER MODES FROM ALL TANKS (A) USING ONLY THE 3 MBS; (B) USING ONLY THE 12 WAIT-MODE COMPONENTS

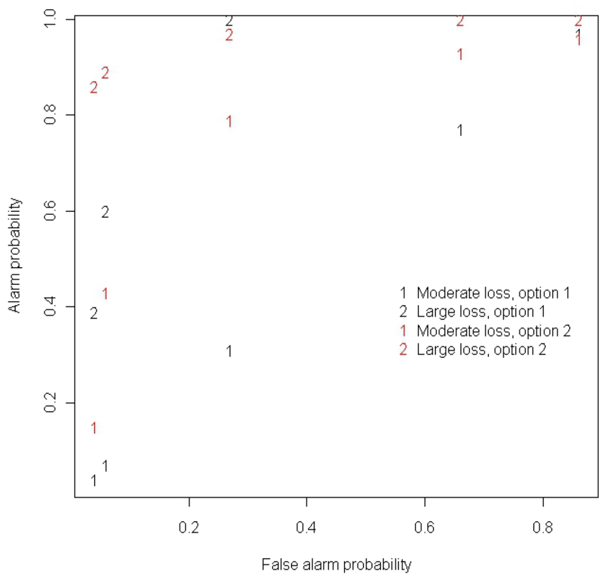


FIG. 15 DETECTION PROBABILITY FOR OPTIONS 1 AND 2. OPTION 1 IS 13 SEPARATE PAGE TESTS. OPTION 2 IS THE MAHALANOBIS DISTANCE FROM MEAN OF ZERO-LOSS DISTRIBUTION. AP RESULTS FOR THE SMALL (NEAR 0.01 TO 0.05) FAPS ARE MOST RELEVANT. DPS FOR THE HIGHER FAP ARE GIVEN FOR COMPLETENESS AND TO EMPHASIZE THAT IT IS IMPORTANT TO CONTROL FOR THE FAP IN "MULTIPLE-TESTING" SITUATIONS

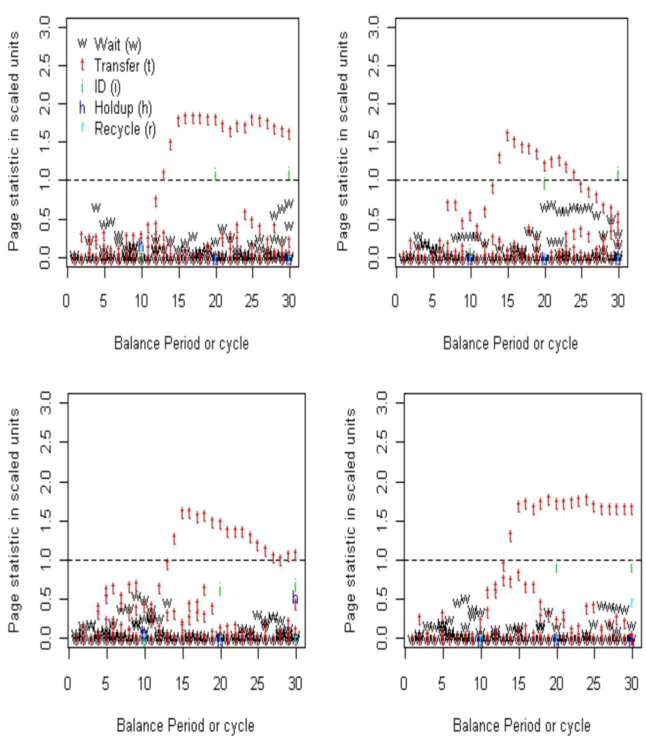


FIG. 16 MODERATE LOSS, 4 REALIZATIONS OF PAGE'S STATISTIC, THE TRANSFER MODES ARE FOR TANK 7 TRANSFERS. NOT ATTEMPTING TO CONTROL OVERALL FALSE ALARM RATE OF 13 NON-INDEPENDENT PAGE STATISTICS

TABLE 1 EXAMPLE DPS FOR THE LOSS OVER 5 WAIT MODES IN TANK 7 AND OVER WAIT AND TRANSFER MODES FROM ALL TANKS. THE SMALL, MODERATE, AND LARGE LOSSES TOTALED 1, 3, AND 30 KG OF PU, RESPECTIVELY. FOR COMPARISON, IF THE LOSS IS 1 SQ = 8 KG SOMETIME DURING BALANCE PERIOD 2 FOR EXAMPLE, THE DPS ARE 0.33, 0.56, AND 0.70, FOR FALSE ALARM PROBABILITIES OF 0.01, 0.05, AND 0.10, RESPECTIVELY

Loss Scenario	False Alarm Probability		
	0.01	0.05	0.10
Loss from Tank 7			
Small	0.02	0.10	0.20
Moderate	0.47	0.83	0.92
Large	1.0	1.0	1.0
Loss from Tank 7, 3 MBs only			
Small	0.02	0.06	0.11
Moderate	0.07	0.30	0.44
Large	1.0	1.0	1.0
Loss from all tanks			
Small	0.01	0.04	0.09
Moderate	0.01	0.06	0.14
Large	0.01	0.11	0.28
Loss from all tanks, 3 MBS only			
Small	0.02	0.07	0.13
Moderate	0.11	0.29	0.39
Large	0.48	0.71	0.79

To summarize Section IX, we considered a 2-tank and a 7-tank example with a few simple loss scenarios. Both period-driven and data-driven pattern recognition for hypothesis testing were numerically illustrated. DPs for additional diversion scenarios are also being estimated using simulation in R. For example, we anticipate high DP for diversion to waste streams that have relatively small predicted amounts of Pu, such as in the waste stream in the 7-tank example. PM and NMA residuals were analyzed on the same statistical footing, following

option 2 in Section V (rather than option 1 which uses a subset of residuals as a master subsystem). We do not claim that any of the candidate diversion detection systems is "optimal," nor have we defined "optimal" in this multivariate sequential testing context. In future work we anticipate tuning the pattern recognition to a few chosen scenarios, plus having a versatile test that has nonzero DP for any diversion, analogous to a multivariate outlier detection test [51]. We also anticipate relying on simulation to estimate alarm thresholds and DPs as in Table 1.

Summary

We have described the safeguards goal to make better quantitative use of PM data, and explored options for continuing to use P(alarm|diversion scenario) as the figure of merit, while using both PM and NMA residuals in the alarm rule. Various alarm rules are being evaluated, all of which involve some type of pattern recognition applied to multivariate time series of PM and NMA residuals that arrive at unequal frequencies.

We believe it is acceptable to tune the pattern recognition to a list of important diversion scenarios to achieve high DP for those scenarios, provided P(alarm|diversion scenario) is non-zero for all scenarios so that the system is at least somewhat robust to any diversion scenario. Esimating P(alarm|diversion scenario) requires modeling and simulating the effects of each diversion scenario, so model uncertainty should be considered in future work. Model uncertainty has been considered in related safeguards contexts [52].

Figure 10 provides the best summary of our strategy, with PM and NMA residuals plotted over 30 days, illustrating that there are several pattern recognition options for developing system alarm rules. We also provide in Appendix 1 a summary flow chart of the analyses.

For the 7-tank example in Section IX, we illustrated pattern recognition options for the period-driven approach, and sequential testing for the data-driven approach. Future work will use a hybrid of data-driven and period-driven approaches.

REFERENCES

[1] Johnson, S., Chesnay, B., Pearsall, C., Takeda, S., Fujumaki, K., Iwamoto， T., "Meeting the safeguards challenges of a commercial reprocessing plant," Proc. 7th International conference on facility operations-safeguards interface, 2004.

[2] Ehinger, B., Chesnay, C. Creusot, J. Damico, S. Johnson, J. Weuster, S. Masuda, M. Kajii, "Solution monitoring applications for the Rokkasho reprocessing plant," 7th International Conference on Facilities Operations-Safeguards Interface, American Nuclear Society. 2004.

[3] Bevan, G., Howell, J., Burr, T., "A tank monitoring evaluation system for a standard three-tank arrangement," Proc. 52nd Annual Meeting Inst. Nuclear Materials Management, 2011.

[4] Burr, T., Bakel, A., Bryan, S., Budlong-Sylvester, K., Damico, J., Demuth, S., Ehinger, M., Garcia, H., Howell, J., Johnson, S., Krebs, J., Myers, K., Orton, C., Thomas, M., "Roles for process monitoring in nuclear safeguards at aqueous reprocessing plants," Journal of Nuclear Materials Management 40(2), 42-53, 2012.

[5] Bakel, A., Burr, T., Demuth, S., Ehinger, M., Frey, K., Garcia, H., Howell, J., Johnson, S., Krebs, J., Orton, C., Schwantes, J., "A dissolver diversion scenario illustrating the value of process monitoring," Proc. 52nd Annual INMM Meeting, 2011.

[6] Orton, C., Fraga, C., Christensen, R., Schwantes, J., "Proof of concept experiments of the multi-isotope process monitor: an online, nondesctructive, near-real-time monitor of spent nuclear fuel reprocessing facilities," Nuclear Instruments and Methods in Physics Research A 672, 38-45, 2011.

[7] Cipiti, B., "Separations and safeguards for performance modeling for advanced reprocessing facility design," Journal of Nuclear Materials Management 39(2), 4-15, 2011.

[8] Bevan, G., Burr, T., Howell, J., "The performance of current solution monitoring evaluation systems approaches in best cases," Proc. 49th Annual Meeting Inst. Nuclear Materials Management, 2009.

[9] Avenhaus, R., Jaech, J., "On subdividing material balances in time and/or space," Journal of Nuclear Materials Management 10(3), 24-33,1981.

[10] Dayem, H., Baker, A., Cobb, D., Hakkila, E., Ostenak, C., “Demonstration of rear-real-time accounting: the AGNS 1980-1981 miniruns,” Los Alamos National Laboratory Report A9942,1984.

[11] Bakel, A., Frey, K., Demuth, S., Johnson, S., Burr, T., Ehinger, M., Garcia, H., Orton, C., Schwantes, J., “Safeguarding the next generation reprocessing plant-solution monitoring in solvent extraction,” Proc. 53rd Annual Meeting Inst. Nuclear Materials Management, 2012.

[12] Garcia, H., Bakel, A., Lin W., Demuth, S., Burr, T., “Evaluating safeguards benefits of process monitoring as compared with nuclear material accountancy,” Proc. 53rd Annual Meeting Inst. of Nuclear Materials Management, 2012.

[13] Garcia, H., Burr, T., Coles, G., Edmunds, T., Garrett, A., Gorensek, M., Hamm, L., Krebs, J., Kress, R., Lambertia, V., Schoenwald, S., Tzanos, C., Ward, R., “Integration of facility modeling capabilities for nuclear nonproliferation analysis,” Progress in Nuclear Energy, 54(1), 96-111, 2011.

[14] Richir, P., Dechamp, L., Janssens-Maenhout, G., Holzleitner, L., Janssens, W., Schwalbach, P., Casteleyn, K., Duinslaeger, L., “Data analysis and interpretation software from solution monitoring towards near real time accountancy,” Proc. 48th Annual Meeting Inst. Nuclear Materials Management, 2007.

[15] Howell, J., Bevan, G., Burr, T., “Inspector interfaces to facilitate subjective reasoning about quantities in trends,” to appear, Computers and Chemical Engineering, 2012.

[16] Venables, W., Ripley, B., Modern Applied Statistics with Splus, Springer: New York, 1999.

[17] Singhal, A., Seborg, D., “Pattern matching in multivariate time series databases using a moving-window approach,” Industrial and Engineering Chemistry Research 41, 3822-3838, 2002.

[18] Kakizawa, J., Shumway, R., Taniguchi, M., “Discrimination and clustering for multivariate time series,” Journal of the American Statistical Association 93(441), 328-340, 1998.

[19] Bleasdale, S., Burr, T., Scovel, J., Strittmatter, R., “Knowledge fusion: an approach to time series model selection followed by pattern recognition,” LA-13095-MS, 1996.

[20] Burr, T., “Statistical Methods in Nuclear Safeguards, Proliferation Resistance, Safeguards and Physical Security,” in Nuclear Engineering Handbook, Springer, D. Cacuci, editor, 2009.

[21] Speed, T., Culpin, D., “The role of statistics in nuclear materials accounting: issues and problems,” Journal of the Royal Statistical Society A, 149(4), 281–313,1986.

[22] Goldman, A., Picard, R., Shipley, J., “Statistical methods for nuclear materials safeguards: an overview,” Technometrics 24(4), 267-275, 1982.

[23] Jones, B., “Near real time materials accountancy using SITMUF and a joint page’s test: improvement of the test,” ESARDA bulletin, 16(5), 1989.

[24] Picard, R., “Sequential analysis of material balances,” Journal of Nuclear Materials Management, 15(2), 38-42, 1987.

[25] Burr, T., Coulter, A., Howell, J., Wangen, L., “Solution monitoring: quantitative and qualitative benefits to safeguards,” Journal. of Nuclear Science and Technology 40(4), 256-263, 2003.

[26] Neuilly, M., Modelling and Estimation of Measurement Errors, Lavoisier Publishing, Paris, France, 1999.

[27] Burr, T., Hemphill, G., “Multi-component Radiation measurement error models,” Applied Radiation and Isotopes 64(3), 379-385, 2006.

[28] Deron, S. et al., “International target values for uncertainty components in fissile isotope and element accountancy for the effective safeguarding of nuclear materials,” Journal of Nuclear Materials Management 30(2), 2002.

[29] Yamaya, K., Nobuharu, S., Iseki, T., Kawai, A., Iwamoto, T., “Plutonium inventory estimation of extraction cycles at Rokkasho reprocessing plant,” Proceedings 50th Annual Meeting Inst.Nuclear Materials Management, 2009.

[30] Ehinger, M., Zack, N., Hakkila, A., Franssen, F., "Use of process monitoring for verifying facility design for large-scale reprocessing plants," LA-12149-MS, ORNL 11856, 1991.

[31] Beyerlein, A., Geldard, J., Weh, R., Eiben, K., Dander, T., "Experimental validation of pulsed column inventory estimators," LA-UR-91-2872, 1991.

[32] Candy, J., Rozsa, R., "Safeguards design for a plutonium concentrator—an applied estimation approach," Automatica, 16, 615-627, 1980.

[33] Burr, T., Suzuki, M., Howell, J., Longo, C., Hamada, M., "Signal estimation and change detection in solution monitoring for safeguards", Nuclear Instruments and Methods in Physics Research a 640, 200–221, 2011.

[34] Howell, J., Scothern, S., "An explicit model-based diagnostic approach in a plutonium nitrate storage facility," Control Engineering Practice 8, 645-656, 2000.

[35] Burr, T., Suzuki, M., Howell, J., Hamada, M., "Loss detection results on simulated tank data modified by realistic effects," Journal of Nuclear Science and Technology 49(2), 209-221, 2012.

[36] Binner, R., Howell, J., Janssens-Maenhous, G., Sellinschegg, D., Zhao K., "Practical issues relating to tank volume determination, process design and control," Industrial Engineering and Chemistry Research 47, 1533-1545, 2008.

[37] R Core Development Team, www.r-project.org, 2004.

[38] Dass, S., Jain, A., "Fingerprint-based recognition," Technometrics 49(3), 262-276, 2007.

[39] Stanbro, W., Sylvester, K., "The role of expert judgment in safeguards,"Journal of Nuclear Materials Management 28(4), 17-20, 2000.

[40] Lyman, E., "Can proliferation risks of nuclear power be made acceptable?" Nuclear Cities Initiative (NCI) 20th conference, Washington DC., 2001 (www.nci.org/conf/lyman).

[41] Burr, T., Hamada, M.S., Orton, C., "Data-driven versus period-driven change detection for process monitoring," Proceedings of the 52nd annual meeting of the Institute of Nuclear Materials Management, 2012.

[42] Crosier, R., "Multivariate generalizations of cumulative sum quality control schemes," Technometrics 30, 291-303, 2008.

[43] Hastie, T., Tibshirini, R., Friedman, J., The Elements of Statistical Learning, Springer, New York, USA, 2001.

[44] Bishop, C., Pattern recognition and machine learning, Springer New York, 2006.

[45] Steinwart, K., Hush, D., Scovel, C., "A classification framework for anomaly detection," Journal of Machine Learning Research 6, 211-232, 2005.

[46] Burr, T., Doak, J., "Distribution-free discriminant analysis," Intelligent Data Analysis 11(6), 651-662, 2007.

[47] Jensen, W., Birch, J., Woodall, W., "High breakdown estimation methods for phase I multivariate control charts," Quality and Reliability Engineering International 23 (5), 615-629, 2007.

[48] Wisnowski, J., Simpson, J., Montgomery, D., "A performance study for multivariate location and shape estimators,"Quality and Reliability Engineering International 18, 117-129, 2002.

[49] Hines, J., Henkel, J., Upadhyaya, B., "Process monitoring techniques for nuclear material diversion detection,"Journal of Nuclear Materials Management 37(3), 4-13, 2009.

[50] Xerri, C. et al., "Control of nuclear material holdup in MOX fuel fabrication plants in Europe," ESARDA Bulletin 31, 69-75, 2002.

[51] Burr, T., Hamada, M. S., Song, J., Wolinsky, M., "Special topics in pattern recognition with applications in nonproliferation," book chapter on pattern recognition, Nova Science, 2011.

[52] Burr, T., Hamada, M.S., "Simultaneous estimation of computer model parameters and model bias," Applied Radiation and Isotopes, 70(8), 1675-1684, 2012.

Appendix 1 Analyses Flow Chart

1) Build a facility model with sufficient fidelity to:

a) predict observables from specified diversion scenarios, and
b) provide model-based predictions of SNM flows to all streams, including waste streams

2) Develop methods to generate PM residuals. Examples:

a) Event marking can be used for solution monitoring to generate PM residuals for each transfer and wait mode for each tank (Section IV).

b) Dissolver cycle time, nitric acid concentration, and temperature can be used to predict SNM in the hull waste which can be compared to neutron-based SNM measurements (Section VIII).

↓

3) Use the facility model to characterize the behavior of the PM and NMA residuals under no-loss and under diversion scenarios as described in Section IX to generate Figures 12-14 from residuals such as shown in Figure 10. Several pattern recognition options are available for this step, including density estimation.

↓

4) To develop a set of system alarm rules, use a hybrid of period-driven and data-driven sequential testing applied to “scores” generated from step (3).

↓

5) Use simulation to estimate P(alarm | diversion scenario) as in Section IX. The alarm rules can include rules learned in step (3) that have high detection probability for specified diversions. In addition, the alarm rules should have reasonable detection probability for unspecified scenarios.